

The equivalence of fluctuation scale dependence and autocorrelations

Duncan J Prindle and Thomas A Trainor

CENPA 354290 University of Washington, Seattle WA 98195 USA

E-mail: prindle@npl.washington.edu, trainor@hausdorf.npl.washington.edu

Abstract. We define optimal per-particle fluctuation and correlation measures, relate fluctuations and correlations through an integral equation and show how to invert that equation to obtain precise autocorrelations from fluctuation scale dependence. We test the precision of the inversion with Monte Carlo data and compare autocorrelations to conditional distributions conventionally used to study high- p_t jet structure.

1. Introduction

Fluctuations in nuclear collisions measured at a single bin size or *scale* could arise from many possible configurations of a multiparticle momentum distribution and are therefore difficult to interpret. However, the *scale dependence* of fluctuations over a significant scale interval does contain detailed information about multiparticle correlations which can be extracted with the proper techniques [1]. Information about the absolute location of event-wise structure is lost, but those aspects depending only on position *difference* are retained in the form of autocorrelation distributions [2]. In this paper we consider fluctuations on binned momentum space, Pearson's correlation coefficient and autocorrelation density ratios for multiplicity n and transverse momentum p_t . Autocorrelations can be inferred directly from pair ratios or from a fluctuation/autocorrelation *integral equation* which we derive. Inverting the integral equation we obtain autocorrelations which can be compared with more conventional conditional distributions.

2. Fluctuations and correlations on binned spaces

Correlation analysis of nuclear collisions reveals information arising from event-wise changes in multiparticle momentum distributions. The data system is an ensemble of event-wise particle distributions on momentum space (p_t, η, ϕ) or (y_t, η, ϕ) , where p_t is transverse momentum, m_t is transverse mass, η is pseudorapidity, ϕ is azimuth and $y_t \equiv \ln\{(m_t + p_t)/m_0\}$ is transverse rapidity with pion mass assigned to m_0 . The momentum space is bounded by a detector acceptance, and the space within the acceptance is binned according to one or more bin sizes. Particle number n is distributed on the full momentum space. It is useful to think of transverse momentum p_t or rapidity y_t as a continuous measure distributed on angle subspace (η, ϕ) and sampled by individual particles. Number n and transverse momentum p_t or rapidity y_t correlations can be considered both separately and in conjunction.

An ensemble of event-wise histograms on momentum subspace x is represented schematically in Fig. 1 (first panel), with bins a , b singled out. Events can be compared with the ensemble-mean

distribution to determine *relative* information, measured by fluctuations of bin contents about their means. The ensemble mean of event-wise pair distributions on space (x_1, x_2) in Fig. 1 (second panel) can be compared to a reference distribution consisting of cartesian products of single-particle mean distributions. The *difference* reveals correlations in the two-particle distribution corresponding to fluctuations in the single-particle distribution [3]. That relation is the basis for the integral equation connecting fluctuations to correlations described below.

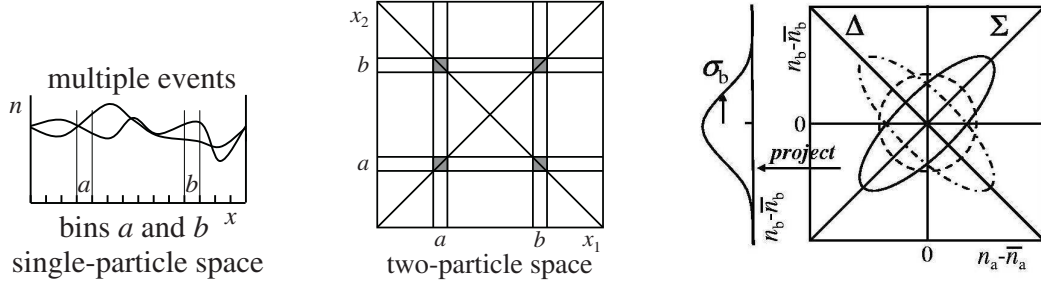


Figure 1. Event-wise distributions on a binned primary space, corresponding two-particle space and possible distributions of bin contents for selected bins a and b in the primary space.

Fig. 1 (third panel) sketches possible frequency distributions on number combinations (n_a, n_b) from bin pair (a, b) in the first panel. The ellipses represent half-maximum contour lines for gaussian-random fluctuations. The three cases correspond to correlation (solid curve), anticorrelation (dash-dot curve) and no correlation (dashed curve) between bins, the last being expected from a mixed-pair or *central limit* reference. The 2D frequency distribution is characterized by two marginal variances and a covariance. The marginal variances for bins a and b are given by $\sigma_{a,b}^2 = \overline{(n - \bar{n})_{a,b}^2} = \overline{n_{a,b}^2} - \bar{n}_{a,b}^2$. The covariance between those bins is given by $\sigma_{ab}^2 = \overline{(n - \bar{n})_a (n - \bar{n})_b} = \overline{n_a n_b} - \bar{n}_a \bar{n}_b = \sigma_\Sigma^2 - \sigma_\Delta^2$, where σ_Σ^2 and σ_Δ^2 are variances along sum and difference diagonals in the third panel. Pearson's correlation coefficient is a measure of *relative* covariance [4]. For bin pair (a, b) it is defined by $r_{ab} \equiv \sigma_{ab}^2 / \sqrt{\sigma_a^2 \sigma_b^2} = \{\sigma_\Sigma^2 - \sigma_\Delta^2\} / \{\sigma_\Sigma^2 + \sigma_\Delta^2\} \in [-1, 1]$. The numerator is the (a, b) covariance and the denominator is the geometric mean of the marginal a and b variances. That coefficient is our basic correlation measure. Quantities r_{ab} , determined for all histogram bin pairs in the second panel, completely represent fluctuations on space x . Variances and covariances depend on the bin size or scale on x . The *scale dependence* of fluctuations is in turn directly related to two-particle correlations, as described in this paper.

3. Object and reference distributions

We distinguish between an object distribution, part of whose correlation content we wish to measure, and a reference distribution which contains by construction information in the object distribution we wish to ignore. We determine object and reference distributions for number n correlations and p_t correlations (the latter require a slightly different treatment, as described below). One single-particle reference for fluctuation measurements is the ensemble-mean histogram on x . Variances and covariances for single bins and bin pairs measure the average information in the event-wise object distribution relative to the ensemble mean.

For two-particle distributions on (x_1, x_2) , object pair density $\rho_{obj}(x_1, x_2)$ is constructed from *sibling* pairs taken from same events, and reference pair density $\rho_{ref}(x_1, x_2)$ is constructed from *mixed* pairs taken from different but similar events. The reference could also be a Cartesian product of single-particle means (therefore *factorizable* by construction). Object and reference distributions are combined in several ways: 1) density of correlated pairs: $\Delta\rho(x_1, x_2) \equiv \rho_{obj}(x_1, x_2) - \rho_{ref}(x_1, x_2)$; 2) density of correlated pairs *per particle pair*:

$\Delta\rho(x_1, x_2)/\rho_{ref}(x_1, x_2) \equiv \rho_{obj}(x_1, x_2)/\rho_{ref}(x_1, x_2) - 1$; 3) density of correlated pairs *per particle*: $\Delta\rho(x_1, x_2)/\sqrt{\rho_{ref}(x_1, x_2)} \equiv \{\rho_{obj}(x_1, x_2) - \rho_{ref}(x_1, x_2)\}/\sqrt{\rho_{ref}(x_1, x_2)}$. 1) and 2) are conventional correlation measures. 3) is unconventional, but closely related to Pearson's correlation coefficient described in the previous section: a relative covariance measure which does not depend on the absolute number of particles *per se*. Pearson's coefficient is ideally suited for testing linear superposition in the context of heavy ion collisions.

4. Autocorrelations

The autocorrelation concept was first introduced to time-series analysis in the form $A(\tau) \equiv 1/T \int_{-T/2}^{T/2} f(t) f(t + \tau) dt$, where τ is the *lag* [2]. The concept is most useful when function $f(t)$ is *stationary*: its correlation structure does not depend on absolute location on time. The information in $f(t)$ is then fully represented by the autocorrelation distribution on τ . The autocorrelation concept can be generalized to spatial correlations. If event-wise structure is randomly positioned on space x then the corresponding ensemble-average two-point distribution on (x_1, x_2) is stationary (not depending on absolute position on x). Distributions in Fig. 2 (left two panels) of measure $\Delta\rho(x_1, x_2)/\rho_{ref}(x_1, x_2)$ on η and ϕ are typical of Au-Au collisions at RHIC [5]. Two-particle distributions can also be defined on sum and difference variables: $\rho(x_1, x_2) \rightarrow \rho(x_\Sigma, x_\Delta)$, with $x_1, x_2 \rightarrow x_\Sigma \equiv x_1 + x_2, x_\Delta \equiv x_1 - x_2$. The data distributions in Fig. 2 exhibit stationarity—they do not depend on sum variable x_Σ —in which case we have $\rho(x_\Sigma, x_\Delta) \rightarrow \rho(x_\Delta)$ to good approximation. We then *average* the two-particle density over x_Σ to obtain the autocorrelation density on x_Δ (still a 2D density, not a projection).

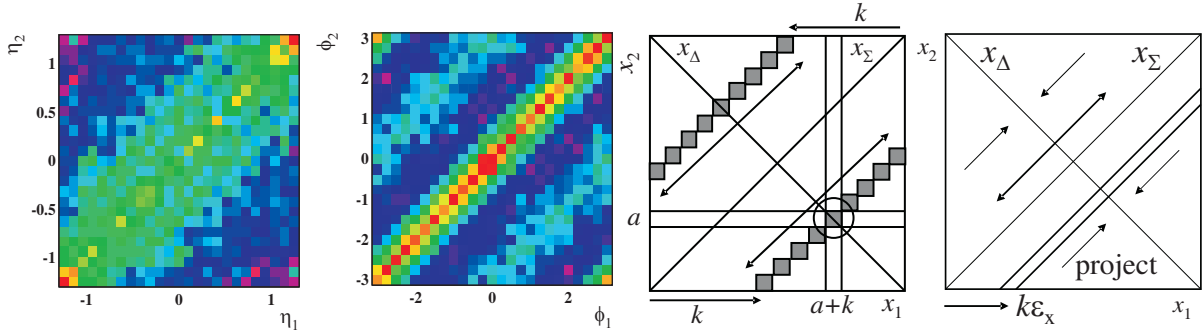


Figure 2. Two-particle correlations on η and ϕ for central Au-Au collisions at 130 GeV, schematic of a binned two-particle space illustrating an autocorrelation average along the k^{th} diagonal and a similar averaging procedure performed directly on difference variable x_Δ .

The autocorrelation on difference variables can be constructed in two ways as shown in Fig. 2 (last two panels): 1) bin space x with *microbins* of size ϵ_x and average the bin contents of space (x_1, x_2) along diagonals, as discussed further below, 2) bin difference variable x_Δ in space (x_1, x_2) directly and form the corresponding pair histogram [1]. In both cases, care must be taken to insure that true averages are obtained and not projections by integration. We define autocorrelation densities $\rho(x_\Delta)$, autocorrelation histograms $A_k(\epsilon_x) \simeq \epsilon_x^2 \rho(2k\epsilon_x)$ and *joint* autocorrelations on two or more difference variables. If the primary distribution on x is truly stationary, the autocorrelation is a *lossless compression* of the two-particle momentum distribution to a lower-dimensional space.

5. Autocorrelation density ratio for number correlations

We now define a universal correlation measure for nuclear collisions. Returning to Pearson's correlation coefficient we make the following approximation

$$r_{ab} \equiv \frac{\sigma_{ab}^2}{\sigma_a^2 \sigma_b^2} = \frac{\overline{(n - \bar{n})_a (n - \bar{n})_b}}{\sqrt{\overline{(n - \bar{n})_a^2} \overline{(n - \bar{n})_b^2}}} \rightarrow \frac{\overline{(n - \bar{n})_a (n - \bar{n})_b}}{\sqrt{\bar{n}_a \bar{n}_b}} = \frac{\bar{n}_a \bar{n}_b - \bar{n}_a \bar{n}_b}{\sqrt{\bar{n}_a \bar{n}_b}}. \quad (1)$$

We replace the marginal number variances in the denominator of r_{ab} by their Poisson values. The result is the histogram equivalent of density ratio $\Delta\rho(x_1, x_2)/\sqrt{\rho_{ref}(x_1, x_2)}$ previously defined. This relative covariance can be interpreted as the number of (anti)correlated pairs per particle (explicit in the last expression). The third combination of object and reference distributions in Sec. 3 and the modified Pearson's normalized covariance in Eq. (1) are equivalent. This density ratio is the basic correlation measure for any bin pair (a, b) on space x . We then define an autocorrelation in terms of density ratios for sets of bins (a, b) on (x_1, x_2) : we combine density ratios as averages along diagonals k , assuming the basic distribution on (x_1, x_2) is approximately stationary. The average on index a along the k^{th} diagonal in Fig. 2 (third panel) is

$$\frac{\Delta A_k(n)}{\sqrt{A_{k,ref}(n)}} \equiv \left\{ \frac{\overline{(n - \bar{n})_a (n - \bar{n})_{a+k}}}{\sqrt{\bar{n}_a \bar{n}_{a+k}}} \right\}_{\bar{a}} \equiv \epsilon_{x_\Delta} \frac{\Delta\rho(n; k\epsilon_{x_\Delta})}{\sqrt{\rho_{ref}(n; k\epsilon_{x_\Delta})}}. \quad (2)$$

This is the autocorrelation definition for analysis of nuclear collisions on angle space (η, ϕ) . This density ratio is an intensive correlation measure which precisely measures relative correlations even for excursions of object and reference densities over orders of magnitude.

6. Extension to transverse momentum p_t correlations

We now extend the definition of the density ratio and its autocorrelation to distributions of transverse momentum p_t on (η, ϕ) . Measurement of $\langle p_t \rangle$ fluctuations is described in [6]. We treat p_t as a continuous measure distributed on space x , with scalar sums of particle p_t in histogram bins. We could write the Pearson's coefficient for transverse momentum by analogy with number correlations in terms of difference $(p_t - \bar{p}_t)$. However, the corresponding per-particle p_t variance can be expressed as the sum of three terms: $\overline{(p_t - \bar{p}_t)^2}/\bar{n} = \overline{(p_t - n\hat{p}_t)^2}/\bar{n} + 2\hat{p}_t \overline{(p_t - n\hat{p}_t)(n - \bar{n})}/\bar{n} + \hat{p}_t^2 \overline{(n - \bar{n})^2}/\bar{n}$, a ' $\langle p_t \rangle$ ' variance, a p_t - n covariance and a number variance. The three terms have forms similar to Pearson's normalized covariance, but with $a = b$ defining corresponding *normalized variances* (also called 'scaled variances'). Each term is important in the overall problem of particle and p_t production. We therefore want to describe the structure of event-wise p_t distributions in terms of $(p_t - n\hat{p}_t)$, independent of but coordinated with the structure of number distributions described in terms of $(n - \bar{n})$. Pearson's correlation coefficient for *transverse momentum* correlations thus takes the form

$$r_{ab} \sim \frac{\overline{(n - \bar{n})_a (n - \bar{n})_b}}{\sqrt{\bar{n}_a \bar{n}_b}} \rightarrow \frac{\overline{(p_t - n\hat{p}_t)_a (p_t - n\hat{p}_t)_b}}{\sigma_{\hat{p}_t}^2 \sqrt{\bar{n}_a \bar{n}_b}}, \quad (3)$$

i.e., a relative p_t covariance as opposed to a relative number covariance. The geometric mean of marginal variances in the denominator is in this case replaced by $\sigma_{\hat{p}_t}^2 \sqrt{\bar{n}_a \bar{n}_b}$, the mean of central-limit expectations for the p_t variances. The factor $\sigma_{\hat{p}_t}^2$ is however omitted in what follows to be consistent with the first term of the per-particle p_t variance expansion above Eq. (3). Factors $\sigma_{\hat{p}_t}^2$ or \hat{p}_t^2 may be introduced as necessary in a subsequent interpretation stage. The corresponding p_t density-ratio autocorrelation is

$$\frac{\Delta A_{kl}(p_t : n)}{\sqrt{A_{kl,ref}(n)}} \equiv \left\{ \frac{\overline{(p_t - n\hat{p}_t)_{ab} (p_t - n\hat{p}_t)_{a+k, b+l}}}{\sqrt{\bar{n}_{ab} \bar{n}_{a+k, b+l}}} \right\}_{\bar{ab}} \equiv \epsilon_{\eta_\Delta} \epsilon_{\phi_\Delta} \frac{\Delta\rho(p_t : n; k\epsilon_{\eta_\Delta}, l\epsilon_{\phi_\Delta})}{\sqrt{\rho_{ref}(n; k\epsilon_{\eta_\Delta}, l\epsilon_{\phi_\Delta})}}. \quad (4)$$

That expression is formulated explicitly in terms of 2D correlations on (η, ϕ) . An analogous 2D expression can be derived for number correlations from Eq. (2).

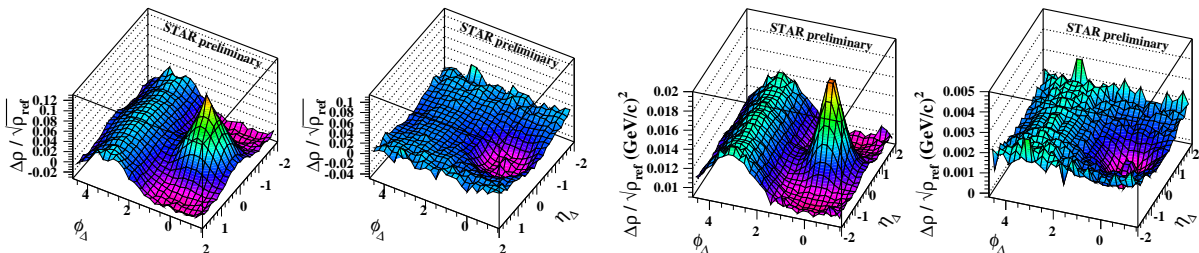


Figure 3. Autocorrelations for 200 GeV p-p collisions determined by direct pair counting: number correlations (left panels) and p_t correlations (right panels) and for charge-independent (left) and charge-dependent or net-charge (right) correlations in each case.

Fig. 3 shows joint autocorrelations of per-particle density ratios on angle difference variables for number correlations (first two panels) and p_t correlations (second two panels) and for charge-independent (like-sign + unlike-sign) pair combinations (left panel of each pair) and charge-dependent (like-sign – unlike-sign) pair combinations (right panel of each pair) [7]. Contributions from self pairs were excluded. These joint autocorrelations, determined directly by pair counting, provide full access to all angular correlations on (η, ϕ) but are computationally expensive for larger event multiplicities. We can also obtain these autocorrelations by inverting fluctuation scale dependence, but with much less computation effort.

7. Relating fluctuations and correlations

Fluctuation scale dependence results from event-wise correlation structure in single-particle distributions which can be extracted by solving an integral equation. We first define differential scale-dependent fluctuation measures based on variance differences for number n and transverse momentum p_t fluctuations, then derive the integral equation connecting fluctuations and correlations. Number variance difference $\Delta\sigma_n^2(\delta x) \equiv (\overline{n(\delta x) - \bar{n}(\delta x)})^2 / \bar{n}(\delta x) - 1$ compares the normalized number variance at bin scale δx to the central-limit or Poisson expectation 1 for that quantity. The difference reflects correlations in the number distribution beyond that of a random distribution of points. The variance difference can be interpreted as the total number of correlated pairs (per-bin number variance σ_n^2) minus the number of Poisson-correlated pairs (self pairs \bar{n}) per particle (divided by \bar{n}). The corresponding variance difference for p_t fluctuations is $\Delta\sigma_{p_t:n}^2(\delta x) \equiv (\overline{(p_t(\delta x) - n(\delta x)\hat{p}_t)^2} / \bar{n}(\delta x) - \sigma_{\hat{p}_t}^2)$. Those per-particle quantities, defined in terms of normalized variances, are consistent in form with Pearson’s normalized covariance.

A fluctuation measurement at a single scale (STAR detector acceptance) is shown in the first panel of Fig. 4 [6]. The frequency histogram on random variable $(p_t - n\hat{p}_t)/(\sqrt{\bar{n}}\sigma_{\hat{p}_t})$ is compared to a central-limit reference (green dotted curve, $\sigma_{\hat{p}_t}$ is the ensemble-average single-particle variance). The p_t variance difference $\Delta\sigma_{p_t:n}^2(\delta x)$ defined previously compares variances of data and reference to reveal a variance excess. Questions then arise how to interpret the fluctuation measurement and how to compare it to measurements made with other detectors.

Fluctuation measurements in different bin-size or scale intervals determine different regions on a *common* distribution representing fluctuation scale dependence, as shown in the second panel of Fig. 4 [9]. The variance difference from the first panel corresponds to the single point at the apex of the surface in the second panel. The surface is obviously structured, but what does the structure mean? Fluctuation scale dependence is the running integral of an autocorrelation. The corresponding integral equation is a linear relation between a variance difference and an

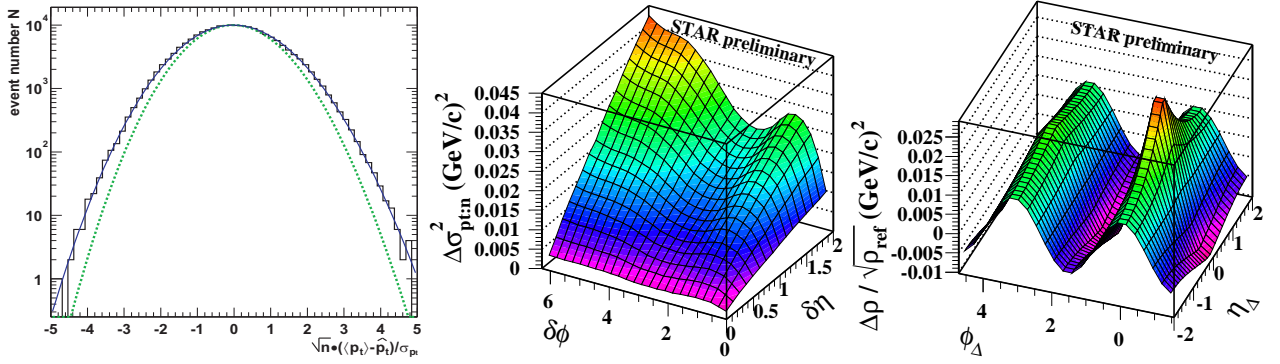


Figure 4. $\langle p_t \rangle$ fluctuations measured at the STAR detector acceptance (histogram) compared to a central-limit reference (dotted curve), the scale dependence of $\langle p_t \rangle$ fluctuations within the STAR acceptance and the corresponding p_t autocorrelation obtained by inversion.

autocorrelation, including a kernel representing the binning scheme. We can express the per-particle variance difference on scales $(\delta\eta, \delta\phi)$ as 2D discrete integral

$$\Delta\sigma_{p_t:n}^2(m\epsilon_\eta, n\epsilon_\phi) = 4 \sum_{k,l=1}^{m,n} \epsilon_\eta \epsilon_\phi K_{mn;kl} \frac{\Delta\rho(p_t : n; k\epsilon_\eta, l\epsilon_\phi)}{\sqrt{\rho_{ref}(n; k\epsilon_\eta, l\epsilon_\phi)}}, \quad (5)$$

with kernel $K_{mn;kl} \equiv (m - k + 1/2)/m \cdot (n - l + 1/2)/n$ representing the 2D macrobin system. This is a Fredholm integral equation which can be inverted by standard numerical methods to obtain autocorrelation density ratio $\Delta\rho_{kl}/\sqrt{\rho_{ref,kl}}$ as a per-particle correlation measure on difference variables $(\eta_\Delta, \phi_\Delta)$ [10]. The third panel of Fig. 4 shows the autocorrelation corresponding to the data in the second panel, with directly interpretable structure [9].

8. Derivation of the integral equation

The derivation relies on coordinating the contents of macrobins, which determine the binning scale for fluctuations, and microbins, which are the basis for the discrete numerical integration of the autocorrelation. We divide the acceptance into macrobins of varying size δx (scale dependence) and microbins of fixed size ϵ_x . The scale-dependent variance is an average over all bins of a particular scale within the acceptance as in Fig. 5 (first panel), where the acceptance is Δx , the macrobin size is δx and the macrobin number is M . The average over macrobins is re-expressed as an average over microbins, as shown in Fig. 5 (second panel). The average over microbins is re-arrange into a sum (over k) of different diagonals and an average over microbins on the k^{th} diagonal. The diagonal average is the k^{th} element of an autocorrelation histogram.

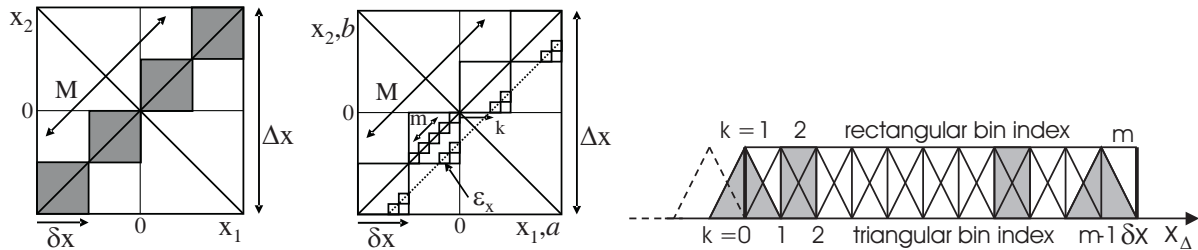


Figure 5. Macrobins of scale δx on a two-particle space after binning a primary space x , microbins of size ϵ_x relative to macrobins and two microbin schemes, one on x , the other on x_Δ .

The p_t variance difference at scale δx in the first line of Eq. (6) is re-expressed as a 2D sum over microbins in the second line, with m microbins in each macrobin. Mean multiplicity $\bar{n}(\delta x)$ in a macrobin relates to the mean microbin multiplicity in that macrobin as $\bar{n}(\delta x) = m \bar{n}(\epsilon_x)$, which relation is applied in the second line. The single-particle variance in the first line corresponds to self pairs which are excluded from subsequent pair sums. The single-particle variance term is consequently dropped. The third line is a rearrangement of the sum over (a, b) into a sum over diagonal index k and a sum over indices a, b subject to the diagonal constraint $a - b = k$,

$$\begin{aligned}
\Delta\sigma_{p_t:n}^2(\delta x) &\equiv \overline{(p_t(\delta x) - n(\delta x) \hat{p}_t)^2} / \bar{n}(\delta x) - \sigma_{\hat{p}_t}^2 & (6) \\
&= \sum_{a,b=1}^m \frac{\{p_t(\epsilon_x) - n(\epsilon_x) \hat{p}_t\}_a \{p_t(\epsilon_x) - n(\epsilon_x) \hat{p}_t\}_b}{m \bar{n}(\epsilon_x)} \\
&= \sum_{k=1-m}^{m-1} K_{m:k} \frac{\bar{n}_k}{\bar{n}} \left[\frac{1}{m - |k|} \sum_{1 \leq a, b \leq m}^{a-b=k} \frac{\sqrt{\bar{n}_a \bar{n}_b}}{\bar{n}_k} \cdot \frac{\{\dots\}_a \{\dots\}_b}{\sqrt{\bar{n}_a \bar{n}_b}} \right] \\
&\equiv \sum_{k=1-m}^{m-1} K_{m:k} \frac{\Delta A_k(p_t : n; \epsilon_x)}{\sqrt{A_{k,ref}(n; \epsilon_x)}} \\
&\rightarrow 2 \sum_{k=1}^m K_{m:k} \frac{\Delta A_k(p_t : n; \epsilon_x)}{\sqrt{A_{k,ref}(n; \epsilon_x)}} \equiv 2 \sum_{k=1}^m \epsilon_x K_{m:k} \frac{\Delta \rho(p_t : n; k \epsilon_x)}{\sqrt{\rho_{ref}(n; k \epsilon_x)}}. & (7)
\end{aligned}$$

Factors have been introduced so that the expression in square brackets is an average along the k^{th} diagonal of normalized microbin covariances, with weighting factor $\sqrt{\bar{n}_a \bar{n}_b} / \bar{n}_k \sim 1$. The fourth line identifies the square bracket as a ratio of autocorrelation histograms. Factor $\bar{n}_k / \bar{n} \sim 1$ has been absorbed into the definition of the autocorrelation ratio, but could be extracted as a correction factor. In the fifth line the binning system has been shifted by $1/2$ bin according to Fig. 5 (third panel), and symmetry about the origin has been invoked (requiring an additional factor 2) to simplify the indexing. The histograms are finally converted to densities by including the microbin width. The last line, generalized to 2D on angle variables (η, ϕ) , is Eq. (5). That integral equation provides computationally cheap $O(n)$ access to autocorrelations.

9. Inversion and regularization

Eq. (5), a discrete Fredholm integral equation, is a matrix equation of the form $\mathbf{D} = \mathbf{T}\mathbf{I} + \mathbf{N}$, with data \mathbf{D} , image \mathbf{I} and statistical noise \mathbf{N} [10]. In principle, one could simply invert the matrix equation to obtain the image. However, \mathbf{T}^{-1} is effectively a differentiation and acts therefore as a high-pass filter in the language of electrical engineering. The $\mathbf{T}^{-1}\mathbf{N}$ statistical noise term strongly dominates the image derived from a simple matrix inversion, and must be substantially reduced by smoothing or ‘regularization’ [11]. The procedure involves treating the image as a matrix of free values in a χ^2 fit subject to *Tikhonov regularization*: minimize $\chi_\alpha^2 \equiv \|\mathbf{D} - \mathbf{T}\mathbf{I}_\alpha\|^2 + \alpha \|\mathbf{L}\mathbf{I}_\alpha\|^2$, including Lagrange multiplier α which controls the role of local gradient operator \mathbf{L} . The first term measures the data-integrated-image mismatch, the second term measures small-wavelength noise on the image. The latter is equivalent to a compensating low-pass filter which offsets the effect of the differentiation. The resulting image is represented by $\mathbf{I}_\alpha = \mathbf{T}_\alpha^{-1}(\mathbf{D} - \mathbf{N}_\alpha)$ which estimates true image \mathbf{I} .

Regularization is illustrated in a power-spectrum context by Fig. 6 (first panel): a tradeoff between information loss and noise suppression. The basis for choosing the optimum α is illustrated in the second panel. $\|\mathbf{D} - \mathbf{T}\mathbf{I}_\alpha\|^2$ (dots) is signal loss and $\|\mathbf{L}\mathbf{I}_\alpha\|^2$ (triangles) is residual noise. As noted, α controls a compensating low-pass filter: small values retain all signal and a large amount of noise from the differentiation, larger values reduce noise, and finally distort the signal by over smoothing. The optimum value is determined as in Fig. 6 (second

panel). This example illustrates that there are clear criteria for choosing an optimum α so that negligible information is lost from the image while statistical noise is greatly attenuated.

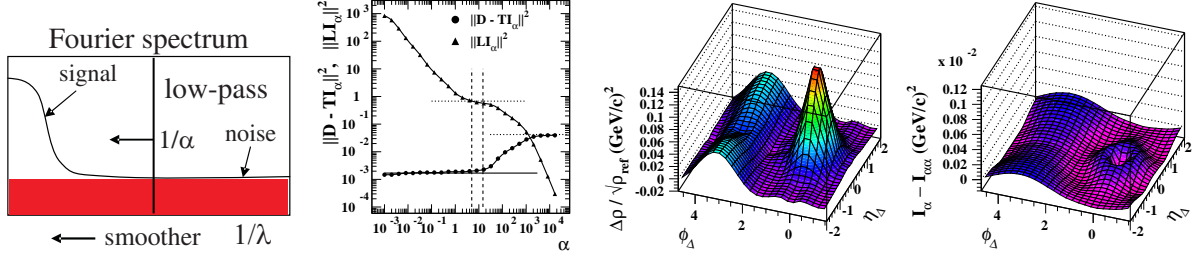


Figure 6. Sketch of Fourier power spectrum, distribution of two terms in χ^2 on Lagrange multiplier α , autocorrelation from fluctuation inversion and corresponding smoothing error.

Statistical and systematic errors are determined by looping through integration and differentiation twice in the sequence $\mathbf{D} \rightarrow \mathbf{I}_{\alpha} \rightarrow \mathbf{D}_{\alpha} \rightarrow \mathbf{I}_{\alpha\alpha}$, including inversion, forward integration and second inversion. Difference $\mathbf{D} - \mathbf{D}_{\alpha}$ estimates statistical error on the data and may itself be inverted to determine residual statistical error on image \mathbf{I}_{α} . Difference $\mathbf{I}_{\alpha} - \mathbf{I}_{\alpha\alpha}$ estimates the smoothing error. Fig. 6 (right two panels) shows image \mathbf{I}_{α} and corresponding smoothing error $\mathbf{I}_{\alpha} - \mathbf{I}_{\alpha\alpha}$ for data from the Hijing Monte Carlo [12], dominated by jet correlations [8].

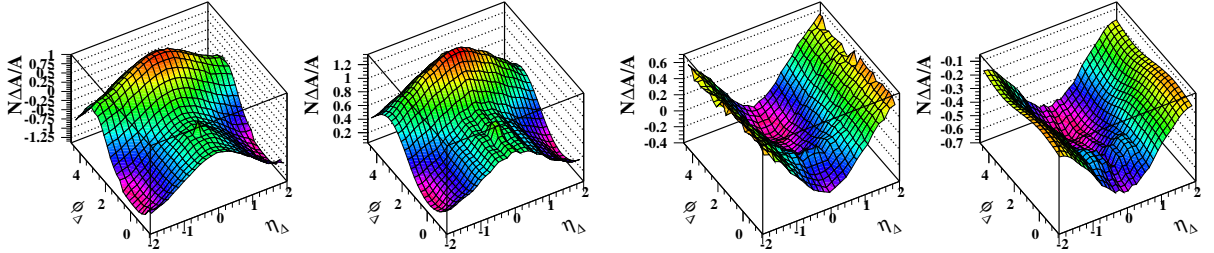


Figure 7. Autocorrelations from the Pythia Monte Carlo for charge-independent (left panels) and charge-dependent (right panels) correlations and direct (left) *vs* inversion (right) methods.

A precision comparison of autocorrelations obtained directly by pair counting and indirectly by fluctuation inversion is presented in Fig. 7. The data were obtained from the Pythia Monte Carlo [13]. The four panels in sequence are charge-independent direct and inverted autocorrelations and charge-dependent direct and inverted autocorrelations. The agreement between methods is excellent.

10. Correlation types

Fig. 8 illustrates types of correlation measurements. The left panels show an autocorrelation obtained by $\langle p_t \rangle$ fluctuation inversion (first panel) and a sketch of corresponding two-bin correlations (second panel) for an arbitrary bin separation. In this type of correlation, structure may occur in most or all events, but with random position on (η, ϕ) . The autocorrelation of a positive-definite measure (such as n or p_t) must be positive-definite at the origin (a variance). Elsewhere (covariances), positive autocorrelation bins indicate correlation (solid curve in second panel), negative bins indicate *anticorrelation* (dash-dot curve), as in the magenta areas adjacent to the positive same-side peak in the first panel. The dashed circle represents the reference.

The right panels illustrate a case where localized structure with approximately fixed position is present in a minority of events. The third panel shows correlations on transverse rapidity

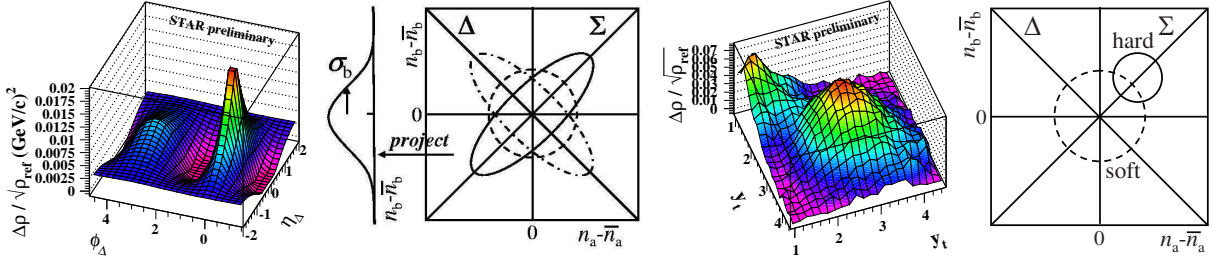


Figure 8. p_t autocorrelation with regions of positive and negative covariance, distributions of bin-pair contents illustrating corresponding correlation (solid) and anticorrelation (dash-dot) trends, distribution on (y_{t1}, y_{t2}) illustrating soft and hard components, and distributions of bin-pair elements illustrating the role of rare hard events in producing positive-definite covariance.

y_t (that distribution is *not* an autocorrelation). In the fourth panel distributions are sketched for two event classes. The dashed curve represents common (*e.g.*, soft) events, the solid circle represents exceptional events (*e.g.*, hard events which contain a detectable parton scatter) from an ensemble of p-p collisions. Positive covariances result from the exceptional events which have an additional multiplicity contribution localized on the y_t space and (in contrast to the autocorrelation example at left) occurring at a nearly fixed position over the event ensemble. The normalized covariance density comparing sibling and mixed pairs in the third panel reveals the contribution from exceptional events (parton fragments).

11. Autocorrelations and conditional distributions

Conventional studies of jet correlations in A-A collisions employ a *leading-particle* analysis (invoked when full jet reconstruction is not possible) [14]. The goal is to estimate a parton momentum by that of the highest- p_t (above some threshold) particle in a collision—the leading or trigger particle. The analysis utilizes two or three *conditional* distributions as illustrated in Fig. 9 (first two panels), where for the sake of comparison transverse momentum p_t has been replaced by transverse rapidity y_t . The condition on (y_{t1}, y_{t2}) is a rectangle representing asymmetric trigger- and associated-particle y_t conditions. Trigger region $(\Omega_{t1}, \Omega_{t2})$ is by construction displaced from the sum diagonal. Angular correlations are also defined as conditional distributions relative to the trigger-particle position in single-particle angle space (η, ϕ) . The angular correlation is plotted (using pseudorapidity as an example) on *conditional* angle difference $\Delta\eta \equiv \eta - \eta_{\text{trigger}}$, which represents an event-wise shift of the *single-particle* angle origin.

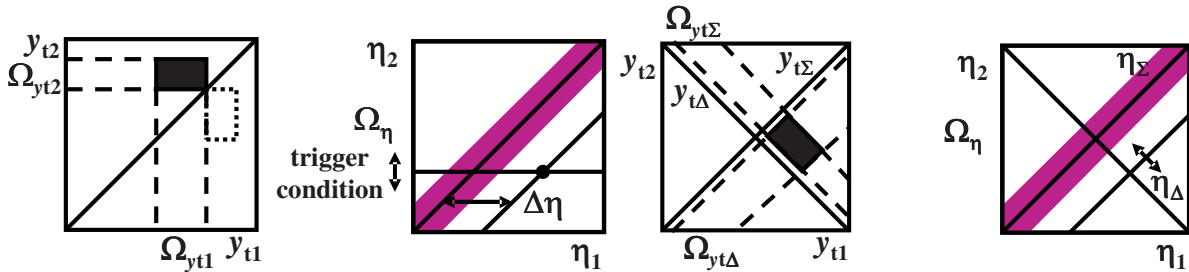


Figure 9. Conditions defined on (y_{t1}, y_{t2}) and (η_1, η_2) for a leading-particle (trigger) analysis compared to a condition on $(y_{t\Sigma}, y_{t\Delta})$ and autocorrelation on η_Δ invoking no leading particle.

Alternatively, one can abandon attempts to estimate parton p_t *per se* and use a technique involving no conditions on momentum, or a symmetric condition on (y_{t1}, y_{t2}) with no conditions

on angle variables, as illustrated in Fig. 9 (second two panels). The fragment distribution on (y_{t1}, y_{t2}) can be studied in its own right. Cut conditions on (y_{t1}, y_{t2}) can be defined to study corresponding changes in jet angular morphology on (η, ϕ) . Angular autocorrelations invoking no trigger condition and defined on symmetric difference variables such as $\eta_\Delta \equiv \eta_1 - \eta_2$ access a minimum-bias parton population. Difference variable η_Δ spans the diagonal axis of the 2D (η_1, η_2) space. Variables $\Delta\eta$ and η_Δ are numerically different, support different distributions and should not be confused (the same comment applies to $\Delta\phi$ and ϕ_Δ). The correlation measures used in the two cases may be compared:

$$\frac{\Delta\rho(n; \eta_\Delta, \phi_\Delta)}{\sqrt{\rho_{ref}(n; \eta_\Delta, \phi_\Delta)}} \quad vs \quad \frac{1}{N_{trig}} \frac{d^2 N_{pair}}{d\Delta\eta d\Delta\phi} \equiv \frac{\rho(n; \Delta\eta, \Delta\phi)}{\int_{\Omega_{trig}} dp_t \rho(n; p_t, trig)}. \quad (8)$$

The latter depends directly on leading- or trigger-particle acceptance Ω_{trig} , and a reference must be provided *a posteriori* by defining a model function to describe the background.

12. Summary

We have derived an integral equation which connects fluctuation scale dependence to corresponding autocorrelations. Inversion of the integral equation reveals autocorrelations which are equivalent to those from pair counting. Fluctuations are thereby interpretable in terms of underlying two-particle correlations. Autocorrelations are complementary to leading-particle techniques for analysis of jet correlations. Definition of normalized variances, covariances and variance differences is tightly constrained by a number of considerations, including linearity, relating fluctuations to correlations, minimizing statistical bias and insuring that correlation structure is maximally accessible and interpretable over a broad range of contexts.

This work was supported in part by the Office of Science of the U.S. DoE under grant DE-FG03-97ER41020.

- [1] Trainor T A, Porter R J and Prindle D J 2005 *J. Phys. G: Nucl. Part. Phys.* **31** 809 Eprint hep-ph/0410182
- [2] Box G E P and Jenkins G 1976 "Time series analysis: forecasting and control" Holden-Day
Bendat J S and Persol A G 1986 "Random data — analysis and measurement procedures" 2nd Edition John Wiley & Sons NY
- [3] Trainor T A 2000 Preprint hep-ph/0001148
- [4] Pearson K (1896) *Phil. Trans. Royal Soc.* **187** 253
- [5] Adams J et al (STAR Collaboration) 2004 Preprint nucl-ph/0406035 submitted to *Phys. Rev. Lett.*
- [6] Adams J et al (STAR Collaboration) 2003 Preprint nucl-ex/0308033 to be published in *Phys. Rev. C*
- [7] Porter R J and Trainor T A (STAR Collaboration) 2005 *Acta Phys. Polon. B* **36** 353
- [8] Liu Q J, Prindle D J and Trainor T A 2004 Preprint hep-ph/0410180 to be published in *Phys. Lett. B*
- [9] ptsca analysis ref: qm04?
- [10] Kirsch A 1996 "An Introduction to the Mathematical Theory of Inverse Problems" Springer-Verlag New York
Groetsch C W 1993 "Inverse Problems in the Mathematical Sciences" Vieweg Publishing Wiesbaden
- [11] Morozov V A 1993 "Regularization Methods for Ill-Posed Problems" CRC Press 1993
Engl H W, Hanke M and Neubauer A 1996 "Regularization of Inverse Problems" Kluwer, Dordrecht
Groetsch C W 1984 "The Theory of Tikhonov Regularization for Fredholm Equations of the first Kind" Pitman Boston 1984
Corbard T *et al* 1998 Preprint astro-ph/9806359
- [12] Wang X N and Gyulassy M 1991 *Phys. Rev. D* **44** 3501
Gyulassy M and Wang X N 1994 *Comp. Phys. Comm.* **83** 307
- [13] Sjöstrand T 1994 *Comput. Phys. Commun.* **82** 74
Sjöstrand T, Lönblad L, Mrenna S and Skands P 2003 Preprint hep-ph/0308153.
- [14] Adler C et al 2003 *Phys. Rev. Lett.* **90** 082302